

# **NOAA/ESRL/GSD/AMB Extensions to WMO RAOB verification**

Bill Moninger, Stan Benjamin, March 2011

## **Background**

Response to a NOAA/ESRL Physical Sciences Review (March 2010) action item

"Actions and due dates: Based on the referenced paper by Moninger et al., GSD will prepare a report on the use of extensions to WMO standards for raob verification by February 8, 2011."

## **1. Introduction**

In 2006 GSD's Assimilation and Modeling Branch (AMB) developed a new RAOB verification procedure for model evaluations. This procedure is discussed in detail in Moninger et al. (2010, *Wea. Forecasting*); this focused discussion is drawn from that paper.

**Under our previous verification procedure**, similar to that used at NCEP and other WMO centers:

- RUC-RAOB comparisons were made only at mandatory sounding levels.
- RAOB data that failed quality control checks in operational models (in our case, the RUC) were not used.
- Also, in our case, verification used RUC data interpolated horizontally and vertically to 40-km pressure-based grids from the RUC native coordinate (isentropic-sigma 20-km) data.

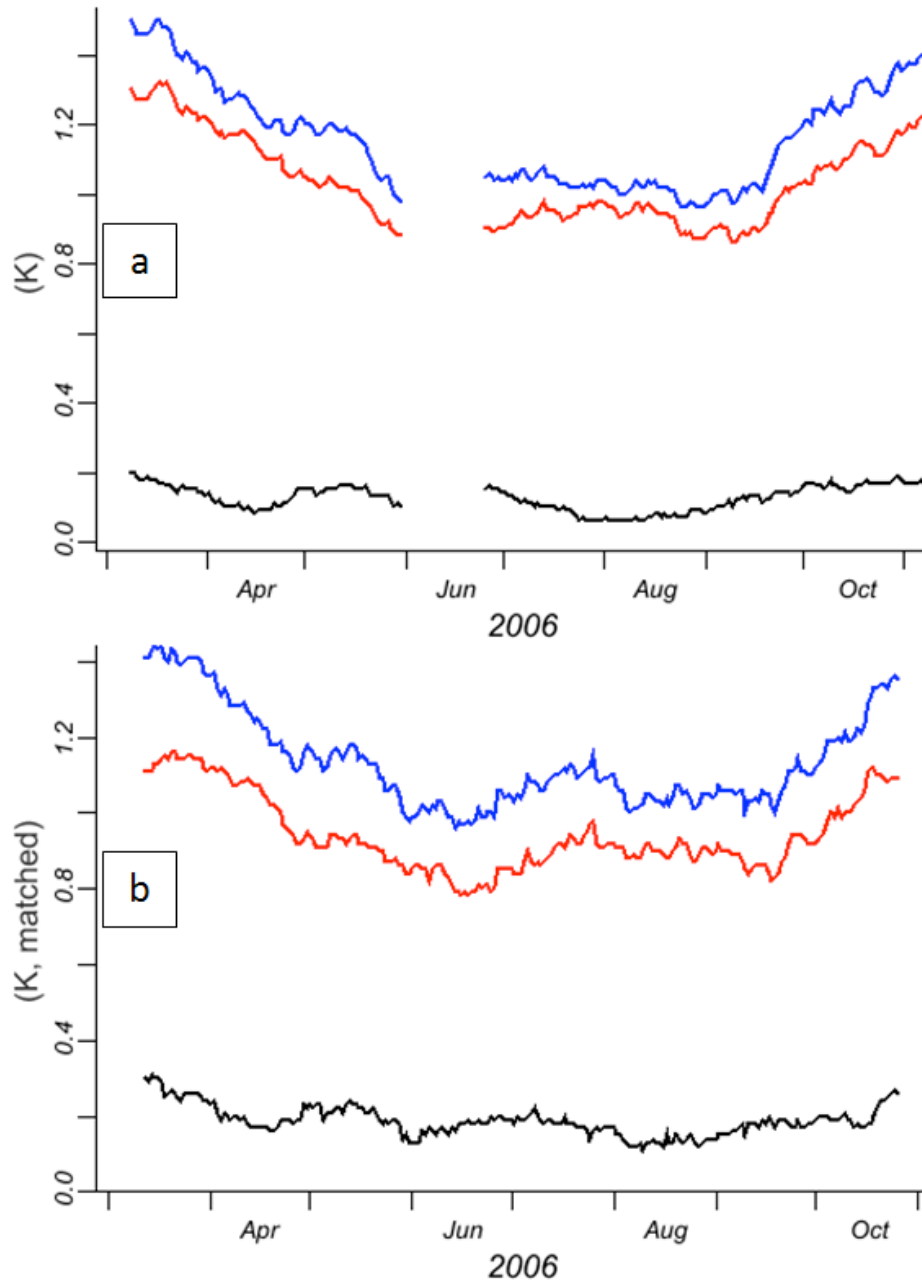
**Under the new verification system:**

- Full RAOB soundings, interpolated to every 10 hPa, are compared with model soundings.
- Model soundings, interpolated to every 10 hPa, are generated directly from native-coordinate files for various models (RUC (isentropic-sigma hybrid), Rapid Refresh (sigma), HRRR (sigma), FIM (isentropic-sigma hybrid)) at various resolutions (regional: 20-km or 13km or 3km (HRRR) resolution; global: 0.5 lat/lon). Additionally, some verification

is performed using data that have already been interpolated to 25-hPa pressure levels (for NAM, GFS, vertically interpolated RUC)

- Comparisons are made every 10 hPa up from the surface.
- No RAOB data are automatically eliminated based on difference from operational RUC analysis data. Obviously erroneous RAOB soundings are periodically eliminated by hand. (Fifteen such RAOB soundings were eliminated between 23 February 2006 and December 2008.)

## **2. Comparison between old and new verification systems**



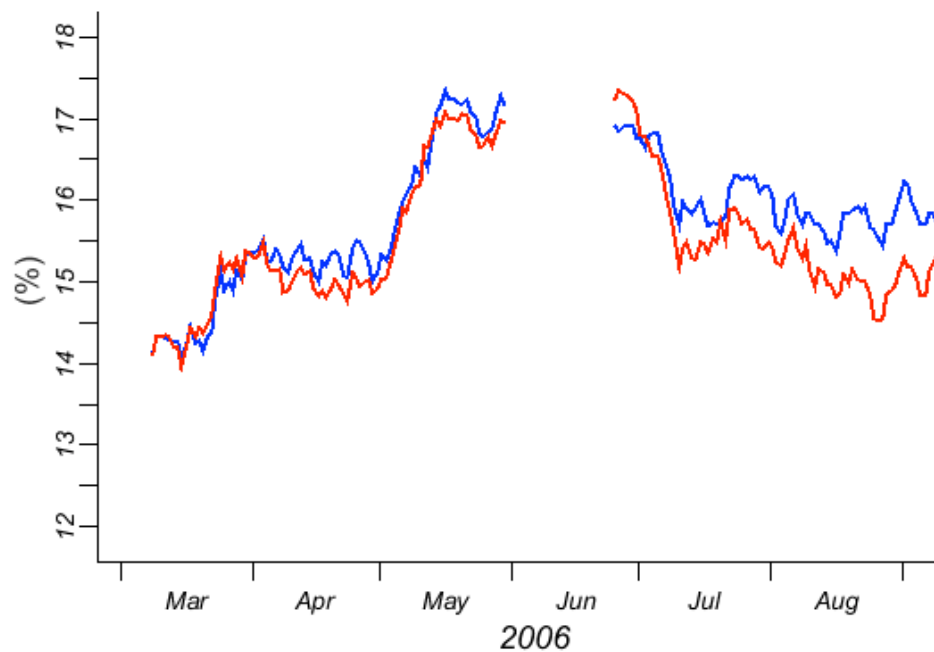
*Fig. 1. 850-hPa temperature 3-h forecast (valid at 0000 UTC, Great Lakes Region) RMS difference between model and RAOBs for 3-h RUC without-TAMDAR (blue) and RUC with-TAMDAR (red), and difference between the two curves (the TAMDAR impact) (black). a) old verification, b) new verification. 30-day running averages.*

To compare the old and new verification methods, we look at the temperature impact at 850 hPa (improvement in skill of RUC forecasts) from of a particular data source called TAMDAR (Moninger et al, 2010). For most of the verified variables at various levels, the old and new verifications give nearly identical answers, as shown in [Figs. 1a,b](#). For this variable and level, the difference in QC screening between the old and new verification made almost no difference. Almost identical results were evident, with an average 0.2 K improvement from assimilating TAMDAR. But not all results were so similar, as we will discuss below.

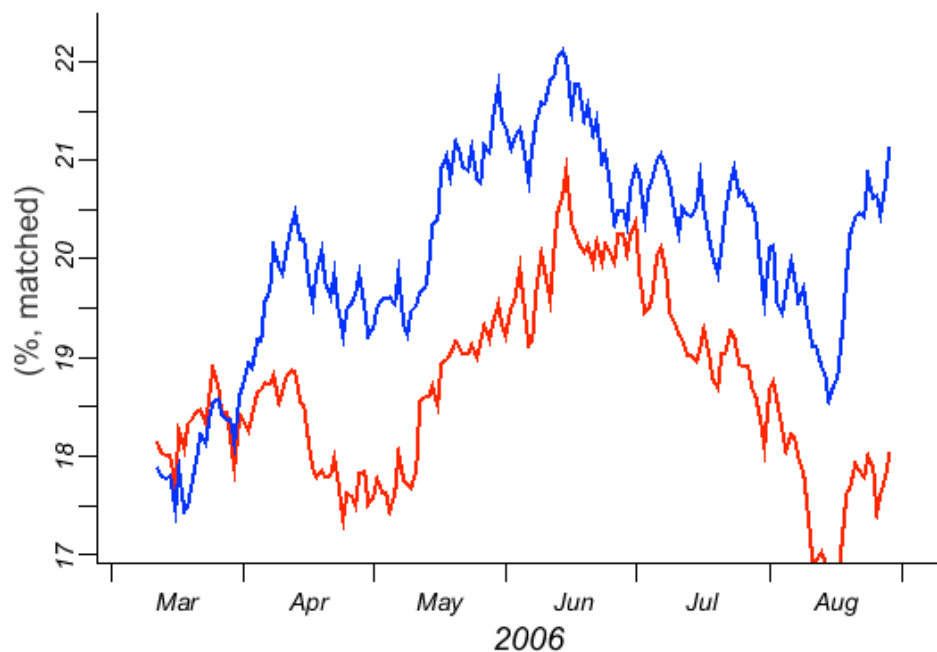
The new verification system has allowed us more vertical precision; we can now, for instance, inspect a data source's impact in the lowest 1500 m above the surface, below 850 hPa. Moreover, inclusion of more RAOB data revealed previously obscured data impacts on relative humidity forecasts. These impacts had previously been obscured because some correct RAOB data were rejected by the old verification system—primarily at 500 hPa—and inclusion of these data changed verification results, especially for RH in the middle troposphere. No longer excluding RAOB data based on their difference from operational RUC values has made a substantial difference in the new verification of 600-400 hPa RH forecasts, as shown in the next example.

A comparison using the old and new verification for 500-hPa 3-h forecast RH RMS error for model runs with and without TAMDAR is presented in [Figs. 2](#) and [3](#). The new verification ([Fig. 3](#)) yields higher RMS error because of the use of **all** RAOB RH values. However, the new verification also shows a much greater *difference* between the TAMDAR and no-TAMDAR runs indicating that the previously missing RAOB data have affected verification of the two cycles unequally. Apparently, assimilation of TAMDAR RH observations improve RUC RH forecasts in cases with large errors in the middle-troposphere where RAOB values were being flagged using the old verification method. Note

that the spacing on the vertical axis is equal, even though the magnitude of the error is larger with the new verification.



*Fig. 2. RMS RH 500 hPa 3-h forecast error for RUC forecasts (blue – without TAMDAR, red – with TAMDAR) against RAOBs for the old verification system (centered at 15% RH). 30-day running averages.*

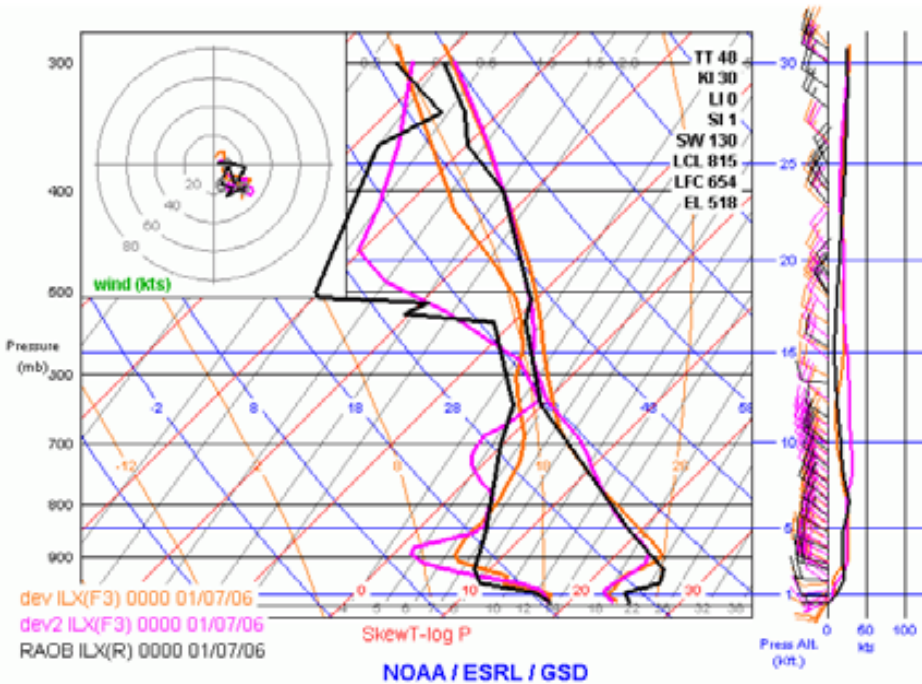


*Fig. 3. As in Fig. 2 but for the new verification system (centered at 20% RH).*

*Table 1. RH values at 500 hPa – 0000 UTC 1 July 2006*

name	RAOB	3h No-TAM	3h TAM	
ILN	33	61	48	
TOP	57	83	75	
PIT	3	76	33	<--
BUF	8	37	7	
OAX	15	53	41	
DTX	14	15	11	
APX	6	6	9	
GRB	30	18	31	
MPX	9	28	33	
ABR	85	90	87	
INL	26	10	21	
DVN	16	39	41	
ILX	19	84	40	<--

To see why this is so, we look at a particular case. [Table 1](#) shows 500-hPa RH values for RAOB observations and the 3-h RUC forecasts, all valid at 0000 UTC 1 July 2006. The old verification did not use the 500-hPa RH RAOBs at PIT (Pittsburgh, PA) and ILX (Lincoln, IL). In both cases (see soundings in Figs. [4](#) and [5](#)), strong subsidence layers were evident, with very dry air with bases just below 500 hPa, accompanied by sharp vertical moisture gradients in the 500-520 hPa layer. The QC screening algorithm used in the previous verification method flagged the 500-hPa RH observations at these two stations since the operational RUC analysis did not maintain this vertical gradient quite as sharply as in the full RAOB data. In both of these cases, the TAMDAR data led the RUC with TAMDAR to better capture this vertical moisture gradient.



*Fig. 4. Soundings at ILX (Lincoln, IL) for 0000 UTC 1 July 2006. RAOB in black, RUC no-TAMDAR 3-h forecast in orange, RUC with-TAMDAR 3-h forecast in magenta.*

[Figure 4](#) shows the observed RAOB and 3-h forecasts for RUC soundings for ILX. The with-TAMDAR forecast sounding (magenta) suggests that TAMDAR had detected a dry layer at 500 hPa. Nearby RAOBs (not shown) also suggest that the observed dry layer at and above 500 hPa was real.

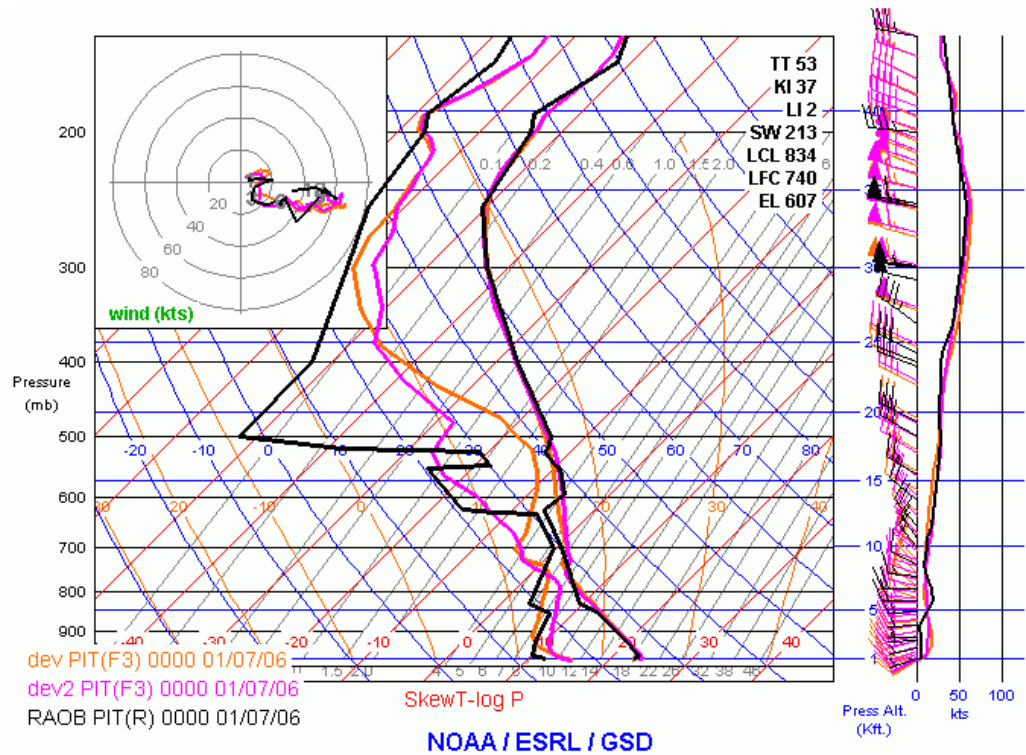
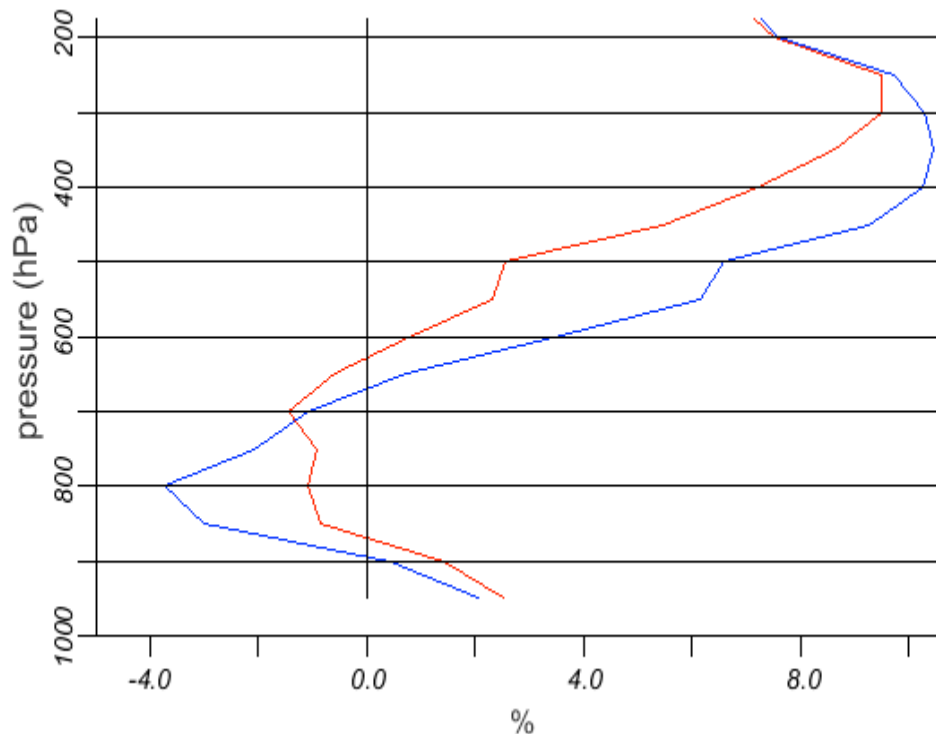


Fig. 5. As for Fig. 4 but for PIT (Pittsburgh, PA).

Figure 5 shows the soundings for PIT. In this case, the accuracy of the dry RAOB observation at 500 hPa is less clear, but is not obviously wrong. Apparently, the much stronger TAMDAR impact shown in Fig. 3 between the no-TAMDAR and with-TAMDAR 500-hPa RH forecasts with the *new* verification screening is attributable to these cases with very sharp vertical moisture gradients near 500 hPa, also suggested by Szoke et al. (2007). Assimilation of the TAMDAR data allows RUC to better capture these features.



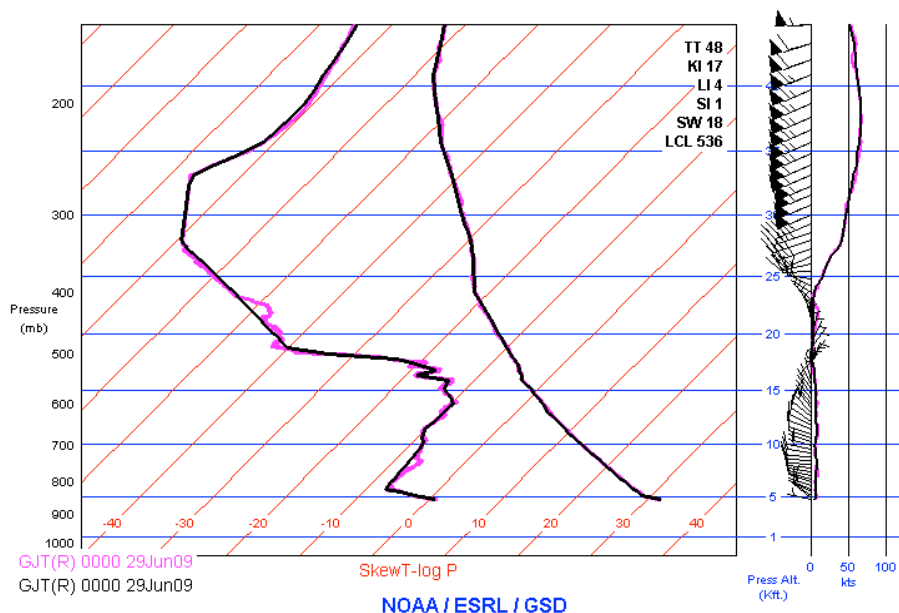
*Fig. 6. Vertical profile for RH bias (model-minus-RAOB) for the no TAMDAR RUC (blue) and with-TAMDAR RUC (red) 3-h forecasts with respect to RAOBs at 00 UTC in the Great Lakes region for the Apr–Aug 2006 period.*

This new verification system also provides much finer vertical resolution than the old, and provides data below 850 hPa. Fig. 6 shows a vertical profile of RH bias for RUC with and without TAMDAR. Note that the RH bias of both models starts positive (more moist than RAOBs) near the surface, becomes negative between approximately 900 and 700 hPa, then becomes increasingly positive with increasing altitude. The old verification system produced data on only three levels at and below 500 hPa (500, 700, and 850 hPa), thereby obscuring vertical variations such as these.

Since some of the finer resolution results from interpolating linearly in log-p between significant levels, we investigated the extent to which this interpolation might differ from actual atmospheric values. One-second resolution data now available from the Radiosonde Replacement System (Facundo, 2004) allowed us to study this. In order to test the effect of interpolation over relatively large pressure ranges, we chose a sounding with relatively few significant levels (Grand Junction on 0000 UTC 29 June 2009, [Fig. 7](#)). In this case, the interpolation extended over pressure ranges up to



120 hPa (between 820 and 700 hPa). For this sounding we calculated the average and RMS difference for temperature, relative humidity, and wind, between the one-second data and the 10-mb interpolated sounding. Results are shown in [Table 2](#) for various pressure bands, and are lower by a factor of 3 to 10 than the RMS differences we typically see in our data deprivation experiments. Thus, we are confident that our interpolation scheme is not obscuring or skewing our forecast impact results and that the linear (in log-p) approximation between RAOB significant levels agrees well with the one-second data, especially for temperature.



*Fig 7. Grand Junction, Colorado RAOB for 0000 UTC 29 June 2009. Data interpolated to 10 hPa from mandatory and significant levels are shown in black; data from one-second data are shown in magenta. (One-second wind barbs are not shown.)*

Table 2. Bias and RMS differences between 10-hPa interpolated and one-second RAOB data from GJT sounding, 0000 UTC 29 Jun 2009

Pressure (hPa)	N/N-RH	T bias/rms (°C)	RH bias/rms (%)	Speed bias / vector wind rms (m/s)
1000 to 800	122	-0.02/0.02	-0.00/0.24	-0.24/0.7

800 to 700	260	-0.02/0.02	-0.52/0.84	0.10/0.93
700 to 600	278	-0.02/0.02	0.37/0.95	-0.44/0.80
600 to 500	338	-0.00/0.01	-0.57/2.52	0.10/1.26
500 to 400	398	0.00/0.01	-0.30/0.95	-1.09/2.85
400 to 300	521	-0.00/0.01	0.09/0.10	0.04/0.87
300 to 200	633	-0.02/0.03	-0.18/0.34	-0.62/1.28
200 to 100	935	0.03/0.06	0.07/0.57	-0.08/2.7
100 to 0	1960	-0.01/0.08	-0.01/0.59	-1.35/3.94

### **3. Differences between our verification system and WMO standards**

Our system has been focused specifically on the models we are developing, which are generally regional rather than global, and generally focus on short-range forecasts (generally less than 24 hours). Our global model, the FIM, uses an icosahedral coordinate system, which is expected to lead to improved forecasts in polar regions. This has led to some differences between our system and WMO standards, which are described below.

#### ***Regions***

The official WMO regions (from [http://www.wmo.int/pages/prog/www/DPS/Publications/WMO\\_485\\_Vol\\_I.pdf](http://www.wmo.int/pages/prog/www/DPS/Publications/WMO_485_Vol_I.pdf)) are

North America	25°N–60°N	50°W–145°W
Europe/North Africa	25°N–70°N	10°W–28°E
Asia	25°N–65°N	60°E–145°E
Australia/New Zealand	10°S–55°S	90°E–180°E
Tropics	20°S–20°N	all longitudes
Northern hemisphere	20°N–90°N	all longitudes

extratropics		
Southern hemisphere extratropics	20°S–90°S	all longitudes

The regions we use are

RUC	Rapid Update Cycle Lambert Conformal Conic domain (CONUS and surrounding regions) see <a href="http://ruc.noaa.gov/">http://ruc.noaa.gov/</a>
RR	Rapid Refresh rotated latitude longitude domain (North America including Alaska) see <a href="http://rapidrefresh.noaa.gov/">http://rapidrefresh.noaa.gov/</a>
Global	Entire globe
Tropics	The official WMO region (20°S–20°N)
Southern hemisphere extratropics	20°S–80°S (all longitudes)
Northern hemisphere extratropics	20°N–80°N (all longitudes)
Arctic	70°N - 90°N (all longitudes)
Antarctic	70°S - 90°S (all longitudes)
Large TAMDAR	38 contiguous RAOBs in the Eastern US
Small TAMDAR region	13 contiguous RAOBs near the US Great Lakes

## ***Variables***

The WMO calls for verifying geopotential height (GPH), temperature, winds. We verify GPH for a few versions of the models. However, we verify relative humidity, which apparently isn't called for in the WMO standard. We find RH helpful because it is reported directly by radiosondes and some aircraft.

## ***Levels***

The standard calls for verification on three mandatory levels: 850, 500, 250 hPa. As described above, our considerably finer vertical resolution has

provided a wealth of information about model behavior that could be revealed if we limited our verification to the WMO standard levels.

### ***Forecast length***

The standard calls for verification for forecasts of 24 h, 48 h, 72 h, 96 h, 120 h, 144 h, 168 h, 192 h, 216 h, 240 h .... Because our primary focus has historically been shorter-range forecasts than this, we generally verify forecasts of 0, 1, 2, 3, 6, 9, 12, 18 h. For our global models, we verify forecasts of 0,12,24,48,72,96,120,144,168,192,216,240 h.

### ***Statistics***

The WMO standard calls for producing statistics of mean error, root mean square error (rmse), vector wind rmse, and trend correlation. GSD/AMB generates all of these except trend correlation, and using the same formulations as shown by the WMO standards.

## **4. Summary**

The GSD/AMB RAOB verification follows WMO standards for RAOB verification except for the exceptions or enhancements noted above in Section 3 (regions, addition of RH, addition of 10-hPa vertical resolution beyond just mandatory levels).

### **References:**

Moninger, W., S. G. Benjamin, B. D. Jamison, T. W. Schlatter, T. L. Smith and E. Szoke (2010), [Evaluation of Regional Aircraft Observations using TAMDAR](#), *WEATHER AND FORECASTING*, 25, 627-645, [10.1175/2009WAF2222321.1](#).

Facundo, J., 2004: Update on the implementation of the National Weather Service's Radiosonde Replacement System. *Eighth Symposium on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface*, Seattle, WA, Amer. Meteor. Soc.

Szoke, E., R. S. Collander, B. D. Jamison, T. L. Smith, S. G. Benjamin, W. R. Moninger, T. W. Schlatter, and B. Schwartz, 2007: Impact of TAMDAR data on RUC short-range forecasts. 22nd Weather Analysis and Forecasting Conf., Park City, Utah, Amer. Meteor. Soc.